

Revista Eletrônica de Sistemas de Informação

ISSN 1677-3071

v. 15, n. 1

jan-abr 2016

doi:10.21529/RESI.2016.1501

Sumário

Editorial

Pietro Cunha Dolci, Alexandre Reis Graeml

Foco na sociedade

TI VERDE: MITO OU REALIDADE NA INDÚSTRIA DIGITAL BRASILEIRA

Ana Cristina de Faria, Márcia da Silva Martins, Luciene Diana Siqueira

LEI DE COTAS: UMA AVALIAÇÃO DAS INSCRIÇÕES NO INSTITUTO FEDERAL DE MINAS GERAIS UTILIZANDO KDD

Márcio Teodoro Dias, Marcello Peixoto Bax

Foco na tecnologia

AN ONTOLOGY OF DISEASES CAUSED BY IMPROPER USE OF PESTICIDES

Sandra Xavier de Macedo, Maria Madalena Dias

Fast Track Semead

ANÁLISE QUANTITATIVA DA FORMAÇÃO DO ESTOQUE DE MÃO DE OBRA QUALIFICADA DE PROFISSIONAIS NA ÁREA DE TI

Maurilio Alves Martins da Costa, Henrique Cordeiro Martins



Este trabalho está licenciado sob uma [Licença Creative Commons Attribution 3.0](http://creativecommons.org/licenses/by/3.0/).

ISSN: 1677-3071

Esta revista é (e sempre foi) eletrônica para ajudar a proteger o meio ambiente, mas, caso deseje imprimir esse artigo, saiba que ele foi editorado com uma fonte mais ecológica, a *Eco Sans*, que gasta menos tinta.

This journal is (and has always been) electronic in order to be more environmentally friendly. Now, it is desktop edited in a single column to be easier to read on the screen. However, if you wish to print this paper, be aware that it uses Eco Sans, a printing font that reduces the amount of required ink.

LEI DE COTAS: UMA AVALIAÇÃO DAS INSCRIÇÕES NO INSTITUTO FEDERAL DE MINAS GERAIS UTILIZANDO KDD

AFFIRMATIVE ACTION LAW: NA ASSESSMENT OF THE REGISTRATIONS AT THE FEDERAL INSTITUTE OF MINAS GERAIS USING KDD

(artigo submetido em agosto de 2014)

Márcio Teodoro Dias

Analista de Tecnologia da Informação do
Instituto Federal de Educação, Ciência e
Tecnologia Minas Gerais
marciodias81@gmail.com

Marcello Peixoto Bax

Professor Titular da Universidade
Federal de Minas Gerais
bax@ufmg.br

ABSTRACT

The article reports a case study to analyze the effectiveness of the places reserved in federal educational institutions by the Brazilian Affirmative Action Law. The study contextualizes the law and its implementation at the Federal Institute of Education, Science and Technology of Minas Gerais (IFMG), and sought to evaluate the actual occupancy of the reserved places; to verify the extent to which people who have used their right to the reserve; and realize what was the percentage of people registered to places without the right to do so. Based on concepts of Knowledge Discovery in Databases, the study describes data extraction and preparation and applied a data-mining tool to the extracted data. Data refer to entries in the selection process and enrollment in IFMG for a period of just over a year. The results are analyzed, showing that a significant proportion of reserved places were not occupied; that many candidates who could register for reserved spaces did not; and that many candidates competed for the reserved places without having the requirements defined by law. Finally, the results suggest the need for greater advertisement of the affirmative action policy and call attention to the importance of verifying the data provided by the candidates who candidate for such special placements.

Key-words: information and knowledge management; Affirmative Action Law; knowledge discovery in databases.

RESUMO

O artigo relata um estudo de caso que analisou a efetividade da reserva de vagas em instituições federais de ensino, criada pela Lei de Cotas. O estudo contextualizou a lei e sua implementação no Instituto Federal de Educação, Ciência e Tecnologia de Minas Gerais (IFMG), e buscou avaliar a real ocupação das vagas reservadas; além de verificar até que ponto as pessoas que têm o direito se inscreveram nas vagas reservadas; e perceber qual foi o índice de pessoas que inscreveram às vagas sem ter o direito. Fundamentado em conceitos da área Descoberta de Conhecimento em Bases de Dados, o estudo realizou a extração e preparação dos dados e aplicou ferramenta de mineração nos dados extraídos. Os dados se referem a inscrições nos processos seletivos e matrículas no IFMG no período de pouco mais de um ano. Em seguida os resultados foram analisados, mostrando que uma parcela significativa das vagas reservadas não foi ocupada; que muitos candidatos que poderiam se inscrever nas vagas reservadas não o fizeram; e que muitos candidatos concorrem às vagas reservadas sem possuir os requisitos exigidos por lei. Finalmente, os resultados sugeriram a necessidade de maior divulgação da política de cotas e alertaram para a importância da verificação das informações prestadas pelos candidatos que concorreram às vagas.

Palavras-chave: gestão da informação e do conhecimento; Lei de Cotas; descoberta de conhecimento em bases de dados.

1 INTRODUÇÃO

Não há dúvida de que o acesso da população à educação de qualidade é um fator essencial para o desenvolvimento econômico de qualquer nação.

Com o objetivo de ampliar o acesso ao ensino técnico e superior às populações com maiores dificuldades de ingressar no ensino público federal, em 2012 a legislação brasileira passou a determinar a reserva de pelo menos 50% das vagas em Instituições Federais de Ensino para estudantes oriundos de escola pública, sendo que parte dessas vagas reservadas é destinada a pessoas de baixa renda e/ou que se consideram pretas, pardas ou indígenas.

Este trabalho tem o objetivo geral de pesquisar a efetividade da aplicação da política de reserva de vagas, por meio de um estudo de caso.

Foram estudados os registros de inscrições e matrículas de uma instituição federal de ensino em um período de pouco mais de um ano, bem como os dados de um questionário socioeconômico preenchido pelo candidato durante a inscrição. A pesquisa teve como objetivos específicos: avaliar a real ocupação das vagas reservadas; verificar até que ponto as pessoas que tem o direito se inscreveram nas vagas reservadas; e qual foi o índice de pessoas que concorreram às vagas sem ter o direito.

Técnicas de Descoberta de Conhecimento em Bases de Dados foram utilizadas para obter informações úteis a partir dos dados brutos extraídos de um sistema informatizado. Pretendeu-se que, essas informações, depois de serem interpretadas, pudessem revelar conhecimentos sobre a efetividade da política de reserva de vagas, e que estes conhecimentos apoiassem a tomada de decisões a respeito dessa política.

2 FUNDAMENTAÇÃO TEÓRICA

2.1 LEI DE COTAS

A reserva de vagas para o ingresso nas instituições federais de ensino passou a acontecer após a publicação da Lei nº 12711 de 29/08/2012, conhecida como Lei de Cotas e de suas regulamentações, feitas no Decreto nº 7824 de 11/10/2012 e na Portaria Normativa MEC nº 18 de 11/10/2012, sendo que está prevista a revisão desse programa no prazo de dez anos a partir da publicação da Lei 12711/2012. A reserva de vagas funciona da seguinte forma:

De todas as vagas ofertadas em instituições federais de ensino, no mínimo 50% devem ser reservadas para estudantes que tenham cursado integralmente o nível anterior em escolas públicas. Ou seja, para as vagas de graduação, a exigência se refere ao Ensino Médio, e para as vagas de nível Técnico ao Ensino Fundamental.

No preenchimento dessas vagas, no mínimo 50% deverão ser reservadas aos estudantes com renda familiar igual ou inferior a um salário mínimo e meio *per capita*.

Além disso, devem ser reservadas vagas para os candidatos autodeclarados pretos, pardos ou indígenas em proporção no mínimo igual à de pretos, pardos e indígenas na população da unidade da Federação onde está instalada a instituição, segundo o último censo do Instituto Brasileiro de Geografia e Estatística (IBGE, 2010). A Figura 1, elaborada pelos autores, mostra essa divisão pelos tipos de cotas:

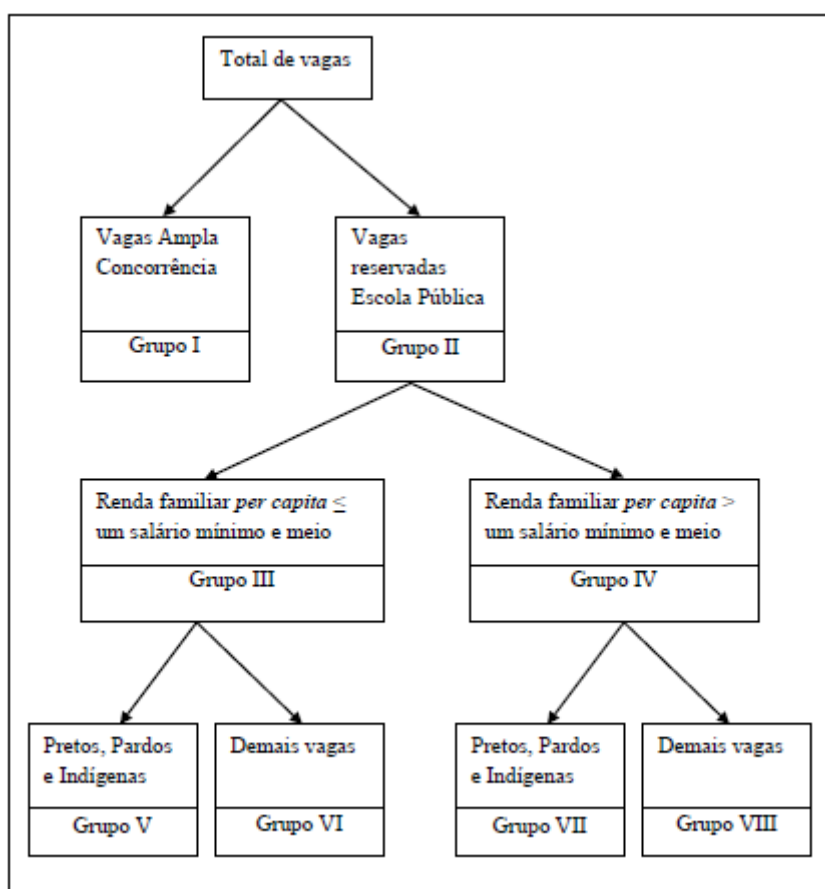


Figura 1. Tipos de cotas
Fonte: elaborada pelos autores

O número de vagas reservadas por curso deve ser informado em Edital do Processo Seletivo de cada Instituição, que tem autonomia para aumentar os percentuais previstos ou mesmo para criar outra modalidade de reserva. Em um primeiro momento, todas as instituições federais de ensino tiveram que implementar a reserva de no mínimo 25% das vagas, tendo 30 de agosto de 2016 como prazo para cumprirem integralmente os critérios definidos na legislação.

O único critério que deve ser comprovado obrigatoriamente é o econômico. Quanto ao estudo em escola pública, cada instituição pode definir se exige ou não essa comprovação. Em relação à cor do candidato,

só é necessária a autodeclaração. Caso após a matrícula seja apurado que o estudante prestou informação falsa, para se beneficiar de uma reserva de vaga a que não tinha direito, essa matrícula deve ser cancelada.

Sempre que o cálculo do número de vagas gerar um valor decimal, é adotado o número inteiro imediatamente superior.

No caso de não preenchimento de vagas reservadas a um determinado grupo, essas devem ser oferecidas a outro grupo de cotas, sendo levadas para a ampla concorrência como última opção.

A participação no sistema de cotas é sempre opcional. Sendo assim, mesmo os candidatos que possuem todos os requisitos para concorrer às vagas reservadas, podem decidir concorrer às vagas destinadas à ampla concorrência, se assim desejarem.

2.2 IFMG

O Instituto Federal de Educação, Ciência e Tecnologia de Minas Gerais (IFMG) é uma autarquia federal, criada pela Lei nº 11.892, sancionada em 29 de dezembro de 2008 pelo então presidente Luiz Inácio Lula da Silva. No artigo 2º desta Lei, os Institutos Federais são definidos como:

Instituições de educação superior, básica e profissional, pluricurriculares e multicampi, especializados na oferta de educação profissional e tecnológica nas diferentes modalidades de ensino, com base na conjugação de conhecimentos técnicos e tecnológicos com as suas práticas pedagógicas, nos termos desta Lei.

O IFMG é hoje formado por dezessete unidades em funcionamento em diferentes cidades do estado de Minas Gerais, sendo que todas são vinculadas a uma Reitoria, em Belo Horizonte. Nessas unidades são oferecidos cursos técnicos e superiores. Quando esta pesquisa foi realizada, foi possível coletar os dados das dez unidades que estavam em funcionamento naquele momento.

O Processo Seletivo Vestibular, forma utilizada para admissão nos cursos, é realizado de forma unificada ao menos duas vezes por ano. Um módulo do sistema (ERP) implantado na Instituição auxilia a gestão do Processo.

No momento da inscrição, além de informar os dados pessoais básicos, e os dados do curso a que deseja concorrer, o candidato deve informar se deseja concorrer a alguma das opções de cota e preencher um questionário socioeconômico, que traz perguntas sobre renda familiar, trabalho, estudos anteriores, entre outras.

O IFMG decidiu implementar, desde o fim de 2012 a reserva de vagas com os percentuais previstos na legislação. Portanto, em todos os processos seletivos acontecidos desde então, foram reservadas 50% das vagas para os alunos oriundos de escolas públicas. De todas as 5231 vagas ofertadas até janeiro de 2014, 2629 foram reservadas para cotas. Considerando que, de acordo com o Censo 2010, o percentual somado de pretos, pardos e indígenas em Minas Gerais é de 53,6% (IBGE, 2010), a

divisão das vagas em cada um dos quatro grupos foi feita como na Tabela 1. A variação do percentual real de vagas ofertadas em cada grupo se explica pela necessidade de arredondamento das vagas ofertadas para cada curso.

Tabela 1. Percentual de vagas reservadas para cada grupo

Candidatos que estudaram em escola pública e têm renda familiar...	... inferior a um salário mínimo e meio <i>per capita</i> e se autodeclararam pretos, pardos ou indígenas (Grupo V)	... inferior a um salário mínimo e meio <i>per capita</i> e não se autodeclararam pretos, pardos ou indígenas (Grupo VI)	... superior a um salário mínimo e meio <i>per capita</i> e se autodeclararam pretos, pardos ou indígenas (Grupo VII)	... superior a um salário mínimo e meio <i>per capita</i> e não se autodeclararam pretos, pardos ou indígenas (Grupo VIII)	Total
Número de vagas reservadas	830	512	792	495	2629
Percentual previsto (sem arredondamento)	13,3%	11,7%	13,3%	11,7%	50,0%
Percentual real (após arredondamento em cada curso)	15,8%	9,8%	15,1%	9,5%	50,2%

Fonte: elaborada pelos autores

2.3 DESCOBERTA DE CONHECIMENTO EM BASES DE DADOS

A expressão em inglês *Knowledge Discovery in Databases* (KDD), geralmente é traduzida em português como Descoberta de Conhecimento em Bases de Dados ou como Descoberta de Conhecimento em Banco de Dados ou como Extração de Conhecimento. É um conceito que define “o processo não trivial de identificar padrões válidos, novos, potencialmente úteis e, finalmente, compreensíveis em dados” (FAYYAD *et al.*, 1996, p. 40-41).

Processo fortemente amparado pela tecnologia da informação, surgiu da necessidade e dificuldade de se explorar grandes volumes de dados disponíveis nas organizações. Segundo Berry e Linoff (1997), o objetivo da KDD é encontrar padrões interessantes ocultos em grandes quantidades de dados, potencializando a habilidade de otimizar a tomada de decisão, pelo uso de métodos automáticos, para aprender com ações passadas.

De acordo com Fayyad *et al.* (1996), existem cinco etapas, mostradas na Figura 2:

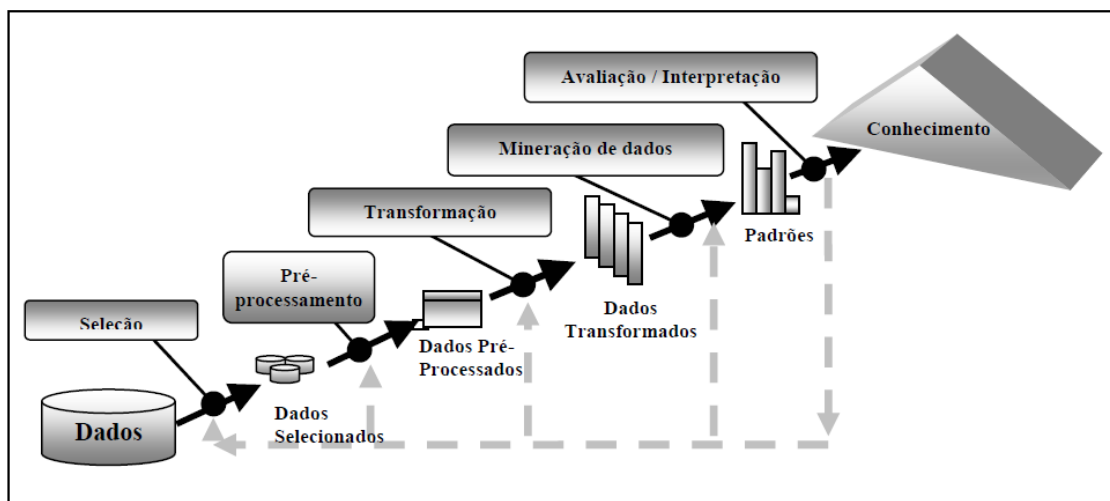


Figura 2. Etapas do Processo de KDD

Fonte: Fayyad *et al.* (1996)

Na seleção são escolhidos e agrupados os dados que serão usados na pesquisa. Esta escolha é feita de acordo com os objetivos que se deseja alcançar. Fayyad *et al.* (1996), afirmam que, diante de cada cenário, o analista utiliza sua experiência anterior, seus conhecimentos e sua intuição para interpretar e combinar os fatos para decidir a estratégia a ser adotada. Para Boente *et al.* (2008, p. 4), esta etapa compreende “a identificação de quais informações, dentre as bases de dados existentes, devem ser efetivamente consideradas durante o processo de KDD”.

O pré-processamento é a etapa de limpeza dos dados, visando melhorar a sua qualidade. Informações ausentes, errôneas ou inconsistentes nas bases de dados devem ser corrigidas de forma a não comprometer a qualidade dos modelos de conhecimento a serem extraídos ao final do processo de KDD (BOENTE *et al.*, 2008).

Na transformação, os dados são codificados em formatos legíveis pelo *software* de mineração, para que possam ser utilizados como entrada dos algoritmos disponíveis. Nessa fase também pode ser obtida mais informação que agregue valor aos registros existentes, vinda de uma consulta a uma base de dados externa, por exemplo (BOENTE *et al.*, 2008).

A mineração de dados, ou *data mining*, é a etapa executada por um *software* específico para este fim. Turban *et al.* (2004) explicam que a mineração de dados recebeu esse nome devido às semelhanças entre a pesquisa de informação de valor em grandes bancos de dados e a garimpagem em busca de minérios preciosos, pois em ambos os processos a quantidade de material a ser garimpado ou sondado é imensa. Segundo Sulaiman e Souza (2000, p. 36) é a “tarefa de estabelecer novos padrões de ‘conhecimento’, geralmente imprevistos, partindo-se de uma massa de dados previamente coletada e preparada para este fim”. Destaca-se nessa definição a importância do trabalho anterior à mineração, de coleta e preparação dos dados.

Por último, há a etapa de avaliação / interpretação, realizada por analistas, que utilizarão o conhecimento adquirido na tomada de decisão. Será tanto mais efetivo quanto maior for o conhecimento do analista. Segundo Santos (2006, p. 88), “o conhecimento descoberto será mais bem utilizado por usuários familiarizados com o conhecimento do domínio trabalhado, pois estes usuários possuem o conhecimento tácito necessário para avaliar as respostas do sistema”.

3 PROCEDIMENTOS METODOLÓGICOS

3.1 PREPARAÇÃO DOS DADOS

De acordo com o objetivo da pesquisa, foi definido que seria necessário selecionar os dados referentes a todos os processos seletivos onde havia reserva de vagas para cotas e que já possuíam as informações referentes à matrícula de todos os candidatos. Quando esta pesquisa foi realizada, havia seis processos seletivos com estas características no banco de dados do ERP do IFMG. Considerando que um deles continha vagas especificamente para EAD (Educação à Distância) e outro para a modalidade PROEJA (Educação de Jovens e Adultos), resolveu-se desconsiderar estes dois processos seletivos, pelo fato de estarem direcionados a um público diferente dos demais. Sendo assim, foi preparada a consulta SQL¹ para extrair em formato CSV² os dados de quatro processos seletivos, com inscrições que aconteceram entre novembro de 2012 e janeiro de 2014.

Os campos extraídos do banco de dados estão descritos no Quadro 1:

Nome	Descrição
unid_curso	Nome do curso precedido pelo nome da unidade do IFMG onde ele é ofertado
sexo	Sexo do candidato no cadastro de pessoa do sistema
cor_raca	Cor/Raça do candidato no cadastro de pessoa do sistema
dt_inscricao	Data em que a inscrição foi efetuada
dt_nascimento	Data de nascimento do candidato no cadastro de pessoa do sistema
status_inscricao	O <i>status</i> da inscrição, com os valores ‘pendente’, ‘cancelada’ ou ‘confirmada’
status_opcao	A situação do candidato no processo seletivo, que vai sendo atualizada ao longo do mesmo.
reserva_vagas_inscrito	Se o candidato selecionou alguma opção de cota, mostra essa opção (EP-Cor-Renda, EP-Cor. EP-Renda ou EP). Se não, é preenchido com AC (Ampla Concorrência)
reserva_vagas_chamado	Campo preenchido com o valor <i>null</i> no momento da inscrição. Após a chamada dos aprovados, ele é atualizado com a opção de cota (EP-Cor-Renda, EP-Cor. EP-Renda ou EP) ou com AC.

¹ *Structured Query Language*

² Comma-separated values

q1	1ª questão do questionário socioeconômico: "Qual o curso para o qual está se inscrevendo?"
q2	2ª questão do questionário socioeconômico: "Qual o seu sexo?"
q3	3ª questão do questionário socioeconômico: "Qual a sua cor/raça?"
q4	4ª questão do questionário socioeconômico: "Qual a sua idade?"
q10	10ª questão do questionário socioeconômico: "Quantos membros de sua família moram em sua casa (incluindo você)?"
q15	15ª questão do questionário socioeconômico: "Qual a renda total de sua família? (Em salários mínimos)"
q19	19ª questão do questionário socioeconômico: "Antes de se matricular no IFMG, até o 5º ano (4ª série), você estudou?" O candidato deveria escolher uma opção entre os seguintes itens: "Sempre em escola pública", "Sempre em escola particular" e "Parte em escola pública, parte em escola particular".
q20	20ª questão do questionário socioeconômico: "Antes de se matricular no IFMG, até o 9º ano (8ª série), você estudou?" O candidato deveria escolher uma opção entre os seguintes itens: "Sempre em escola pública", "Sempre em escola particular" e "Parte em escola pública, parte em escola particular".
q21	21ª questão do questionário socioeconômico: "Antes de se matricular no IFMG, até o 3º ano ensino médio, você estudou?" O candidato deveria escolher uma opção entre os seguintes itens: "Sempre em escola pública", "Sempre em escola particular" e "Parte em escola pública, parte em escola particular".

Quadro 1. Campos extraídos do banco de dados

Fonte: Banco de dados da pesquisa

Observou-se que alguns campos estavam duplicados no cadastro de pessoa, gerado pelo formulário de inscrição e no questionário socioeconômico. No cadastro de pessoa os dados são mais confiáveis, pois estes dados podem ser conferidos pelo candidato durante o processo seletivo, no comprovante de inscrição, e no caso dos alunos que chegaram a se matricular, todos os dados foram conferidos pelo setor de Registro e Controle Acadêmico de cada unidade do IFMG no momento da matrícula. Os dados duplicados do questionário socioeconômico foram também extraídos, com o objetivo de avaliar a sua similaridade com os dados do cadastro de pessoa, como forma de validação do preenchimento do questionário socioeconômico.

Foram extraídos 36542 registros. Após a exclusão das inscrições que não foram confirmadas e de inscrições que estavam duplicadas, chegou-se a um total de 19368 registros a serem pesquisados.

A partir dos dados extraídos, foram gerados novos campos, conforme o Quadro 2, para obter melhores resultados na mineração.

Nome	Descrição
unidade	O nome da unidade do IFMG, obtido pelo campo <code>unid_curso</code> que foi extraído do banco de dados
curso	O nome da curso, obtido pelo campo <code>unid_curso</code> que foi extraído do banco de dados
modalidade	Preenchido com algum dos seguintes valores: Superior-Bacharelado, Superior-Licenciatura, Superior-Tecnologia, Técnico-Integrado ou Técnico-Subsequente
nível	Preenchido com Superior ou Técnico
idade	Calculada com os valores dos campos <code>dt_inscricao</code> e <code>dt_nascimento</code>
faixa_idade	Com o resultado do campo <code>idade</code> , foram criadas as mesmas faixas das respostas do questionário socioeconômico
matriculado	Armazena o valor “S” caso o candidato tenha sido matriculado, que é uma das opções do campo <code>status_opcao</code>
inscrito_em_cota	Armazena o valor “S” caso o candidato tenha sido inscrito em alguma das opções de cota, obtido por meio do campo <code>reserva_vagas_inscrito</code>
chamado_para_cota	Armazena o valor “S” caso o candidato tenha sido chamado a se matricular em alguma das opções de cota, obtido por meio do campo <code>reserva_vagas_chamado</code>
inscrito_cota_cor	Com base no conteúdo do campo <code>reserva_vagas_inscrito</code> , mostra se o candidato se inscreveu para uma opção de cota que contém o critério de cor
inscrito_cota_renda	Com base no conteúdo do campo <code>reserva_vagas_inscrito</code> , mostra se o candidato se inscreveu para uma opção de cota que contém o critério de renda
sempre_escola_publica_qs	Para os candidatos que se inscreveram para um curso de nível superior, calculado pelo conteúdo do campo <code>q21</code> . Para os candidatos que se inscreveram para um curso de nível técnico, calculado pelo conteúdo dos campos <code>q19</code> e <code>q20</code>
cor_ppi_qs	Com base no campo <code>q3</code> , armazena “S” se o candidato tiver se declarado preto, pardo ou indígena
renda_menos_1.5_qs	Os valores dos campos <code>q15</code> e <code>q10</code> possibilitaram o cálculo da renda familiar <i>per capita</i> do candidato para verificar se ela é ou não inferior a um salário mínimo e meio. O campo foi preenchido com “D” onde não foi possível fazer o cálculo (quando o candidato declarou ter renda de mais de oito salários mínimos e mais de cinco membros da família residindo na mesma casa) e com “S” ou “N” nas demais situações.

Quadro 2. Novos campos gerados a partir dos campos extraídos do banco de dados

Fonte: Base de dados da pesquisa

A seguir, foi feita uma verificação de dados não preenchidos ou inconsistentes, sendo possível perceber que algumas datas de nascimento estavam preenchidas com valores improváveis. Foi definido que nas situações em que as idades calculadas fossem abaixo de 13 anos ou acima de 70 anos a faixa de idade seria preenchida com o valor “Desconhecida”.

Para a maioria dos testes, foi utilizado o arquivo com os dados dos 19368 candidatos inscritos. Para a verificação dos candidatos matriculados em cotas, foi gerado um outro arquivo, filtrando apenas os 4579 matriculados no período avaliado.

3.2 MINERAÇÃO E RESULTADOS

Para efetuar a mineração dos dados, foi utilizado o *software* Weka (*Waikato Environment for Knowledge Analysis*) na versão 3.6. O código desse sistema começou a ser desenvolvido em 1993, na Universidade de Waikato, na Nova Zelândia. Construído com a linguagem de programação Java, aceita várias extensões de arquivos como entrada, entre elas a extensão csv.

O algoritmo utilizado foi o *apriori*, apresentado por Agrawal e Srikant (1994). Este algoritmo de associação foi citado por Wu *et al.* (2008) como um dos dez algoritmos de *data mining* mais utilizados no mundo.

3.2.1 Validação dos dados do questionário socioeconômico

Como a pesquisa propõe a utilização de dados do questionário socioeconômico, julgou-se adequado fazer um teste da confiabilidade desses dados, já que estes dados são preenchidos pelo candidato no momento da inscrição e não são conferidos ou alterados posteriormente. Respostas incorretas no questionário podem indicar falta de interesse do candidato no seu preenchimento e conseqüentemente pouca confiabilidade.

Os testes foram feitos apenas com os atributos que se repetem no questionário e no cadastro de pessoa. Segue na Tabela 2 a síntese dos resultados.

Tabela 2. Percentual de similaridade entre campos informados em diferentes locais do sistema

Campo comparado	Similaridade entre os campos pesquisados
modalidade	91%
sexo	99%
faixa idade	94%
cor_raca	90%

Fonte: elaborada pelos autores

O atributo cor/raça não está no formulário de inscrição, mas é solicitado no momento da matrícula. Por isso o teste acima foi feito com os dados dos 4579 candidatos que foram matriculados.

Pelos testes feitos, conclui-se que há um alto nível de confiabilidade nas respostas do questionário socioeconômico, já que em mais de 90% dos casos, a informação foi similar àquela que foi conferida posteriormente durante o processo seletivo ou no momento da matrícula.

3.2.2 Ocupação das vagas reservadas para cotas

Para verificar se as vagas reservadas dentro do período estudado foram efetivamente ocupadas, foi feito o teste com os atributos *reserva_vagas_inscrito* e *reserva_vagas_chamado*. A Tabela 3 mostra o resumo dos testes realizados com os dados de todos os 19368 inscritos e com os 4579 matriculados.

Tabela 3. Inscrições e matrículas nas vagas reservadas para alunos que estudaram em escola pública

1º critério	Todos que estudaram em escola pública (Grupo II)
Percentual de vagas reservadas	50,2%
Número de vagas reservadas	2629
Percentual de inscritos para as vagas reservadas	40,6%
Número de inscritos para as vagas reservadas	7878
Percentual de matriculados nas vagas reservadas	45,1%
Número de matriculados nas vagas reservadas	2065
Vagas reservadas que não foram ocupadas	564

Fonte: elaborada pelos autores

Qualquer candidato que tenha estudado integralmente em escola pública no nível anterior ao que está concorrendo agora, tem o direito de se inscrever em alguma das quatro opções de cotas. Sendo assim, a Tabela 3 mostra resultados gerais, sem especificar o tipo de cota. Por este resultado percebe-se que nem todas as vagas reservadas foram ocupadas, pois apenas 45,1% dos matriculados foram chamados para alguma das opções de cotas. 564 vagas que inicialmente seriam reservadas para alguma opção de cota acabaram sendo ocupadas pela ampla concorrência. Outro ponto importante é que apenas 40,6% dos candidatos decidiram concorrer a alguma das vagas reservadas, sendo este percentual muito menor que a reserva de vagas de 50,2%.

O critério seguinte leva em consideração a renda familiar *per capita*, que deve ser igual ou inferior a um salário mínimo e meio. Nos testes feitos a seguir, foram agrupadas as duas categorias que levam em consideração a remuneração (EP-Renda-Cor e EP-Renda) e as outras duas que não consideram a renda familiar (EP-Cor e EP). A Tabela 4 mostra o resumo dos resultados desse teste.

Pelos resultados percebe-se que nas cotas com critério econômico a ocupação das vagas reservadas foi muito alta, já que os matriculados foram 24,6%. Porém nas demais opções a ocupação foi bem menor, sendo de 20,5%. O percentual de inscritos para quem estudou em escola pública e tem renda familiar *per capita* acima de um salário mínimo e meio (Grupo IV) foi de apenas 15,2%, sendo que 24,6% das vagas estavam reservadas para este público.

Tabela 4. Inscrições e matrículas nas vagas reservadas para alunos que estudaram em escola pública, agrupados pela renda familiar *per capita*

2º critério Todos que estudaram em escola pública e tem renda familiar <i>per capita</i> inferior a um salário mínimo e meio (Grupo III)	... e tem renda familiar <i>per capita</i> superior a um salário mínimo e meio (Grupo IV)
Percentual de vagas reservadas	25,6%	24,6%
Número de vagas reservadas	1342	1287
Percentual de inscritos para as vagas reservadas	25,4%	15,2%
Número de inscritos para as vagas reservadas	4932	2946
Percentual de matriculados nas vagas reservadas	24,6%	20,5%
Número de matriculados nas vagas reservadas	1126	939
Vagas reservadas que não foram ocupadas	216	348

Fonte: elaborada pelos autores

O 3º critério leva em consideração a autodeclaração de cor como preto, pardo ou indígena. Na Tabela 5 estão os resultados da pesquisa com o agrupamento das cotas que levam em consideração a cor (EP-Cor-Renda e EP-Cor) e as cotas que não a levam em consideração (EP-Renda e EP).

Tabela 5. Inscrições e matrículas nas vagas reservadas para alunos que estudaram em escola pública, agrupados pela autodeclaração de cor

3º critério Todos que estudaram em escola pública e se autodeclararam pretos, pardos ou indígenas (Grupos V e VII)	... e não se autodeclararam pretos, pardos ou indígenas (Grupos VI e VIII)
Percentual de vagas reservadas	31,0%	19,2%
Número de vagas reservadas	1622	1007
Percentual de inscritos para as vagas reservadas	28,2%	12,4%
Número de inscritos para as vagas reservadas	5477	2401
Percentual de matriculados nas vagas reservadas	29,5%	15,6%
Número de matriculados nas vagas reservadas	1353	712
Vagas reservadas que não foram ocupadas	269	295

Fonte: elaborada pelos autores

Por causa do arredondamento, o percentual de vagas para os Grupos V e VII foi de 31,0%, bem acima dos 26,6% que seriam ofertadas para estes grupos se o percentual de 53,6% de pretos, pardos e indígenas fosse aplicado no número total de vagas reservadas. De acordo com os resultados, as vagas reservadas para os critérios relacionados à cor estão mais próximos de ser ocupadas. Já em relação às cotas que não levam a cor em consideração, tanto as inscrições quanto as matrículas foram muito menores do que seria esperado. Enquanto 28,2% dos candidatos se inscreveram para concorrer às vagas reservadas aos pretos, pardos e indígenas, apenas 12,4% se inscreveram para as cotas que não levavam em conta a cor.

Para compreender melhor essa diferença, o teste foi realizado para cada tipo de cota. O resumo dos resultados está na Tabela 6.

Tabela 6. Inscrições e matrículas nas vagas reservadas para cada tipo de cota

3º critério Todos que estudaram em escola pública e têm renda familiar <i>per capita</i> inferior a um salário mínimo e meio e se autodeclararam pretos, pardos ou indígenas (Grupo V)	... inferior a um salário mínimo e meio e não se autodeclararam pretos, pardos ou indígenas (Grupo VI)	... superior a um salário mínimo e meio e se autodeclararam pretos, pardos ou indígenas (Grupo VII)	... superior a um salário mínimo e meio e não se autodeclararam pretos, pardos ou indígenas (Grupo VIII)
Percentual de vagas reservadas	15,8%	9,8%	15,1%	9,5%
Número de vagas reservadas	830	512	792	495
Percentual de inscritos para as vagas reservadas	18,5%	6,9%	9,7%	5,5%
Número de inscritos para as vagas reservadas	3590	1342	1887	1059
Percentual de matriculados nas vagas reservadas	16,1%	8,5%	13,5%	7,0%
Número de matriculados nas vagas reservadas	737	389	616	323
Vagas reservadas que não foram ocupadas	93	123	176	172

Fonte: elaborada pelos autores

Pode-se perceber que no Grupo V, a ocupação foi quase total, já que sobraram apenas 93 vagas, de um total de 830 ofertadas. Nos demais grupos, com menos vagas oferecidas, o número de vagas não ocupadas foi consideravelmente maior. Percebe-se também que o Grupo V foi o único em que o percentual de inscrições em cotas foi superior ao percentual de vagas reservadas para o grupo. O Grupo VIII teve o menor percentual de inscrições, o menor percentual de matriculados e, conseqüentemente, foi o que mais cedeu vagas para a ampla concorrência, pois das 495 vagas reservadas, 172 não foram ocupadas, o que representa 34,7%.

3.2.3 Candidatos que poderiam concorrer às vagas mas não o fizeram

Para descobrir até que ponto os candidatos fizeram uso do seu direito à inscrição em cotas, foram testados os atributos que permitem comparar as inscrições em cotas com os dados do candidato que são requisitados para cada tipo de vaga reservada.

Tabela 7. Candidatos que poderiam ter se inscrito em alguma opção de cota

Requisitos	Total de candidatos que possuíam os requisitos	Número de candidatos que se inscreveram nas vagas reservadas, que possuíam os requisitos	Percentual de candidatos que se inscreveram nas vagas reservadas, que possuíam os requisitos
Escola pública	15525	7554	48,7%

Fonte: elaborada pelos autores

Esse resultado mostra que entre todos os inscritos, 15525 poderiam ter se inscrito em alguma das opções de cotas, uma vez que estudaram em escola pública. Desses, apenas 7554 se inscreveram efetivamente, o que significa apenas 48,7% do público potencial de cotistas.

Na Tabela 8 segue um resumo dos resultados agrupando os candidatos em cotas que levavam em consideração os critérios racial e econômico.

Tabela 8. Candidatos que poderiam ter se inscrito em cotas com critérios raciais e econômicos

Requisitos	Total de candidatos que possuíam os requisitos	Entre todos os que possuíam os requisitos, número de candidatos que se inscreveram nas vagas reservadas	Entre todos os que possuíam os requisitos, percentual de candidatos que se inscreveram nas vagas reservadas
Escola pública e cor	10131	5098	50,3%
Escola pública e renda	14290	4648	32,5%

Fonte: elaborada pelos autores

Entre todos os inscritos, 10131 atendiam os critérios para se inscrever nas vagas reservadas a alunos de escola pública que se declararam pretos, pardos ou indígenas (grupos V e VII), mas apenas 50,3% realmente

fizeram a inscrição fazendo prevalecer esse direito. Em relação às cotas para alunos de escola pública com renda familiar *per capita* de até um salário mínimo e meio (Grupo III), dos 14290 que tinham os requisitos, apenas 32,5% concorreram a essas vagas reservadas.

O resultado do teste para cada um dos quatro tipos de cotas está descrito na Tabela 9. Para este teste foram excluídos os 23 candidatos para os quais não foi possível calcular a renda familiar *per capita*:

Tabela 9. Candidatos que poderiam ter se inscrito em cada grupo de cotas

Grupo	Total de candidatos que possuíam os requisitos esperados para cada grupo	Entre todos os que possuíam os requisitos, número de candidatos que se inscreveram nas vagas reservadas	Entre todos os que possuíam os requisitos, percentual de candidatos que se inscreveram nas vagas reservadas
Grupo V	9448	3294	34,9%
Grupo VI	4842	1120	23,1%
Grupo VII	669	258	39,0%
Grupo VIII	543	154	28,3%
Total	15	4826	31,1%

Fonte: elaborada pelos autores

A Tabela 9 mostra o total de candidatos que teria se inscrito em cada grupo se todos os candidatos que possuem o direito tivessem se inscrito na opção correspondente ao seu perfil. Sendo assim, o total de 543 candidatos do Grupo VIII não se refere a todos os alunos que estudaram em escola pública, mas a todos os alunos que estudaram em escola pública, tem renda familiar *per capita* superior a um salário mínimo e meio e não se autodeclararam pretos, pardos ou indígenas. A soma de inscrições nos quatro grupos, que é de 4826, não se refere ao total de inscritos em cotas, que foi de 7878, mas ao total de candidatos inscritos na cota mais adequada ao seu perfil, de acordo com os dados informados. Isso significa que 3052 candidatos (38,7% dos inscritos em cotas), escolheram um grupo diferente do que se esperaria, como, por exemplo, candidatos que declararam ser pretos, pardos ou indígenas no questionário socioeconômico, mas concorreram a uma opção de cota que não leva a cor em consideração.

3.2.4 Candidatos que concorreram às vagas sem ter os requisitos necessários

Foram realizados testes para descobrir inconsistências entre a opção de cotas feita pelos candidatos e o preenchimento do questionário socioeconômico. O algoritmo *apriori* foi executado com os atributos *reserva_vagas_inscrito*, *sempre_escola_publica_qs*, *renda_menos_1.5_qs* e *cor_ppi_qs*. Primeiramente foi avaliada a inscrição de uma maneira geral, sem especificar o tipo de cota, conforme a Tabela 10.

Tabela 10. Candidatos que se inscreveram em cotas possuindo os requisitos

Grupo	Inscritos nas vagas reservadas, cotas	Número de candidatos que possuíam os requisitos	Percentual de candidatos que possuíam os requisitos
Grupo II	7878	7554	96,0%

Fonte: elaborada pelos autores

O requisito comum a todas as opções de cota é ter estudado em escola pública (Grupo II). Seria esperado que 100% dos inscritos em cotas declarassem ter estudado sempre em escola pública, mas 4% deles declararam no questionário socioeconômico que estudaram em escola particular em algum momento.

O teste seguinte foi feito com o agrupamento das cotas de cor e renda, cujos resultados estão na Tabela 11:

Tabela 11. Candidatos que poderiam ter se inscrito em cada grupo de cotas

Grupo	Inscritos nas vagas reservadas, cotas	Número de candidatos que possuíam os requisitos	Percentual de candidatos que possuíam os requisitos
Grupos V e VII	5477	5098	93,1%
Grupos V e VI	4932	4648	94,2%

Fonte: elaborada pelos autores

Nas cotas reservadas para estudantes de escola pública que se declararam pretos, pardos e indígenas (Grupos V e VII), 93,1% possuíam todos os requisitos, de acordo com o questionário socioeconômico. Nas cotas reservadas a quem sempre estudou em escola pública e tem renda familiar *per capita* até um salário mínimo e meio (Grupos V e VI) o percentual foi de 94,2%.

A Tabela 12 mostra o resultado do mesmo teste feito para cada uma das opções de cota.

Tabela 12. Candidatos que poderiam ter se inscrito em cada grupo de cotas

Grupo	Inscritos nas vagas reservadas, cotas	Número de candidatos que possuíam os requisitos	Percentual de candidatos que possuíam os requisitos
Grupo V	3590	3294	91,8%
Grupo VI	1342	1243	92,6%
Grupo VII	1887	1744	92,4%
Grupo VIII	1059	1010	95,3%

Fonte: elaborada pelos autores

Estes resultados indicam que a grande maioria dos candidatos que se inscreve nas cotas possui os requisitos exigidos pela legislação. Porém, mesmo que o percentual não chame a atenção, os 587 candidatos que se inscreveram a vagas reservadas sem ter os requisitos no período estudado representam um número considerável.

4 ANÁLISE DOS RESULTADOS

Para este trabalho, foram executados testes na base de dados preparada para este fim, em busca de informações sobre três aspectos: a ocupação das vagas reservadas, o volume de inscrições nas cotas dos candidatos que cumprem os requisitos e o volume de inscrições nas cotas de candidatos que não cumprem os requisitos (de forma intencional ou não).

4.1 OCUPAÇÃO DAS VAGAS RESERVADAS PARA COTAS

Como mostrado na Tabela 3, 564 vagas reservadas para cotas não foram ocupadas, o que representa 21,4% das vagas reservadas nos processos seletivos estudados.

Observando cada tipo de cota, pela Tabela 6, conclui-se que no Grupo V quase 90% das vagas reservadas foram ocupadas. Esse resultado reflete a procura por estas vagas, que foi a maior entre todos os grupos.

Por outro lado, no Grupo VIII, houve um alto percentual de vagas não ocupadas (34,7%), o que espelha o baixo percentual de inscrições para estas vagas.

Para aumentar o percentual de ocupação das vagas reservadas, chegando aos 50%, seria recomendável aumentar a divulgação sobre os direitos dos candidatos, inclusive na própria interface do sistema que gerencia as inscrições.

4.2 INSCRIÇÕES QUE CUMPREM OS REQUISITOS NECESSÁRIOS ÀS COTAS

Como visto na seção anterior, vagas não foram preenchidas, entretanto candidatos com perfil de cotas (que cumprem os requisitos) se inscreveram ao processo seletivo no Grupo I.

Foram usadas as respostas do questionário socioeconômico para avaliar se o público de cotas em potencial está utilizando seu direito de concorrer às vagas que estão reservadas.

Pelos dados das Tabelas 7 e 8, é possível perceber que menos de 50% dos candidatos que podem concorrer a alguma cota estão fazendo uso desse direito, e que existe uma grande diferença entre os grupos que levam em consideração a cor (50,3%) e os que levam em consideração a renda familiar (32,5%). A Tabela 9 mostra que muitos estão se inscrevendo para uma cota a que eles podem concorrer, mas que não seria a mais indicada, de acordo com seu perfil.

Sugere-se mais uma vez um trabalho de divulgação das cotas, e uma melhor exibição dessas informações no próprio formulário eletrônico de inscrição, como forma de aumentar as inscrições para as vagas reservadas e ajudar o candidato a escolher a opção mais adequada para o seu perfil.

4.3 INSCRIÇÕES QUE NÃO CUMPREM OS REQUISITOS NECESSÁRIOS ÀS COTAS

Os dados encontrados apontam indícios de que há candidatos concorrendo a vagas reservadas sem atender os requisitos, e que não há diferença significativa entre os grupos, pois em todos parece haver candidatos se inscrevendo sem ter direito.

Uma sugestão é que o sistema faça um cruzamento dos dados do questionário com a opção de cota selecionada, apontando a inconsistência, para que os candidatos que marcaram uma opção errada possam corrigi-la antes de finalizar a inscrição.

Outro ponto de atenção é quanto à comprovação da renda familiar e do estudo em escola pública. Pode ser que muitos dos candidatos tenham se candidatado a uma vaga sabendo que não têm o requisito exigido, mas pensando em omitir essa informação no momento da matrícula. Sugere-se um cuidado redobrado para esta comprovação quando da chamada desses candidatos para a matrícula na Instituição.

5 CONSIDERAÇÕES FINAIS

O estudo apontou indícios de que muitas pessoas que teriam o direito a se candidatar às vagas reservadas não o fizeram, o que sugere a necessidade de maior divulgação do programa de cotas.

Por outro lado, parte dos que se inscreveram aparentemente não teriam o direito, o que reforça a necessidade de maior controle. A legislação prevê a obrigatoriedade da comprovação da renda familiar, porém esta é uma comprovação difícil, uma vez que os candidatos podem simplesmente omitir informações para parecer que a sua renda familiar *per capita* é inferior a um salário mínimo e meio.

Quanto ao estudo em escola pública, a legislação permite que a instituição exija a comprovação. O resultado deste trabalho aponta que essa é uma prática recomendável, uma vez que há candidatos que estudaram em escola particular em algum momento ao longo da formação de nível imediatamente anterior e que, mesmo assim, se inscrevem para concorrer às vagas reservadas, sem ter esse direito.

O uso de KDD se mostrou útil para a descoberta das informações, uma vez que foram reveladas situações que não correspondiam ao esperado. A análise do conhecimento gerado pode ser relevante para a tomada de decisão política.

Sugere-se como trabalhos futuros que essa mesma análise dos dados seja feita em outras instituições federais, tanto Institutos quanto universidades e que sejam comparados o nível de ingresso de candidatos que concorreram às vagas reservadas com os que possuíam os requisitos, mas optaram pela ampla concorrência.

REFERÊNCIAS

- AGRAWAL, R.; SRIKANT, R. Fast algorithms for mining association rules in large databases. In: INTERNATIONAL CONFERENCE ON VERY LARGE DATA BASES, 20., 1994, Santiago, Chile. *Proceedings...*, San Francisco: Morgan Kaufmann Publishers Inc., 1994. p. 487-499.
- BERRY, M. J. A.; LINOFF, G. *Data mining techniques – for marketing, sales, and customer support*. John Wiley & Sons, New York, 1997.
- BOENTE, A. N. P.; GOLDSCHMIDT, R. R.; ESTRELA, V. V. Uma Metodologia de Suporte ao Processo de Descoberta de Conhecimento em Bases de Dados. In: V Simpósio de Excelência em Gestão e Tecnologia, 2008, Resende – RJ. V SEGeT, 2008. v. 1.
- BRASIL. Decreto nº 7.824, de 11 de outubro de 2012. Disponível em: <http://www.planalto.gov.br/ccivil_03/_ato2011-2014/2012/Decreto/D7824.htm>. Acesso em: 10 jun. 2014.
- BRASIL. Lei nº 11.892, de 29 de dezembro de 2008. Disponível em: <http://www.planalto.gov.br/ccivil_03/_Ato2007-2010/2008/Lei/L11892.htm>. Acesso em: 10 jun. 2014.
- BRASIL. Lei nº 12.711, de 29 de agosto de 2012. Disponível em: <http://www.planalto.gov.br/ccivil_03/_ato2011-2014/2012/lei/l12711.htm>. Acesso em: 10 jun. 2014.
- FAYYAD, U.; PIATETSKY-SHAPIO, G.; SMYTH, P. From Data mining to Knowledge Discovery in Databases. *AI Magazine*, v. 17, n. 3, p. 37-54, 1996.
- IBGE – Instituto Brasileiro de Geografia e Estatística. Censo demográfico 2010: Características da População e dos Domicílios – Resultados do Universo. Rio de Janeiro, 2011. Disponível em: <http://biblioteca.ibge.gov.br/d_detalhes.php?id=793>. Acesso em: 07 jul. 2014.
- SANTOS, R. S. Aplicação de um modelo preditivo de mineração de dados para apoio à decisão de crédito. 2006. 92 f. Dissertação (Mestrado em Ciência da Informação) – Escola de Ciência da Informação, Universidade Federal de Minas Gerais, Belo Horizonte, 2006.
- SULAIMAN, A.; SOUZA, J. M. Panorama da mineração de dados no Brasil. In: ENCONTRO NACIONAL DE PESQUISA EM CIÊNCIA DA INFORMAÇÃO, 4., 2000, Brasília. *Anais...* Brasília: UnB, 2000.
- TURBAN, E.; LEIDNER, D.; MCLEAN, E.; WETHERBE, J. *Tecnologia da Informação para Gestão: Transformando os negócios na Economia Digital*. 3.ed. Porto Alegre: Bookman, 2004.
- WU, X. *et al.* Top 10 algorithms in data mining. *Knowledge and Information Systems*. Londres. vol. 14, n. 1, p. 1-37. Jan. 2008.